# 8

# Guided Search 4.0

## Current Progress With a Model of Visual Search

*Jeremy M. Wolfe*

Guided search (GS) is a model of human visual search performance, specifically of search tasks in which an observer looks for a target object among some number of distracting items. Classically, models have described two mechanisms of search: *serial* and *parallel* (Egeth, 1966). In serial search, attention is directed to one item at a time, allowing each item to be classified as a target or a distractor in turn (Sternberg, 1966). Parallel models propose that all (or many) items are processed at the same time. A decision about target presence is based on the output of this processing (Neisser, 1963). GS evolved out of the two-stage architecture of models like Treisman's feature integration theory (FIT; Treisman & Gelade, 1980). FIT proposed a parallel, preattentive first stage and a serial second stage controlled by visual selective attention. Search tasks could be divided into those performed by the first stage in parallel and those requiring serial processing. Much of the data comes from experiments measuring reaction time (RT) as a function of set size. The RT is the time required to respond that a target is present or absent. Treisman proposed that there was a limited set of attributes (e.g. color, size, motion) that could be processed in parallel, across the whole visual field (Treisman, 1985, 1986; Treisman & Gormican, 1988). These produced RTs that were essentially independent of the set size. Thus, slopes of RT × set size functions were near zero.

In FIT, targets defined by two or more attributes required the serial deployment of attention. The critical difference between preattentive search tasks and serial tasks was that the serial tasks required a serial "binding" step (Treisman, 1996; von der Malsburg, 1981). One piece of brain might analyze the color of an object. Another might analyze its orientation. Binding is the act of linking those bits of information into a single representation of an object—an object file (Kahneman, Treisman, & Gibbs, 1992). Tasks requiring serial deployment of attention from one item to the next produce RT × set size functions with slopes markedly greater than zero (typically, about 20–30 ms/item for target-present trials and a bit more than twice that for target-absent).

The original GS model had a preattentive stage and an attentive stage, much like FIT. The core of GS was the claim that information from the first stage could be used to guide deployments of selective attention in the second (Cave & Wolfe, 1990; Wolfe, Cave, & Franzel, 1989). Thus, if observers searched for a red letter *T* among distracting red and black letters, preattentive color processes could guide the deployment of attention to red letters, even if no front-end process could distinguish a *T* from an *L* (Egeth, Virzi, & Garbart, 1984). This first version of GS (GS1) argued that *all* search tasks required that attention be directed to the target item. The differences in task performance depended on the differences in the quality of guidance. In a simple feature search (e.g., a search for red among green), attention would be directed toward the red target before it was deployed to any distractors, regardless of the set size. This would produce RTs that were independent of set size. In contrast, there are other tasks where no preattentive information, beyond information about the presence of items in the field, is useful in guiding attention. In these tasks, as noted, search is inefficient. RTs increase with set size at a rate of 20–30 ms/item on target-present trials and a bit more than twice that on the target-absent trials (Wolfe, 1998). Examples include searching for a 2 among mirror-reversed 2s (5s) or searching for rotated *T*s among rotated *L*s. GS1 argued that the target is found when it is sampled, at random, from the set of all items.

Tasks where guidance is possible (e.g., search for conjunctions of basic features) tend to have intermediate slopes (Nakayama & Silverman, 1986; Quinlan & Humphreys, 1987; Treisman & Sato, 1990; Zohary, Hochstein, & Hillman, 1988). In GS1, this was modeled as a bias in the sampling of items. Because it had the correct features, the target was likely to be picked earlier than if it had been picked by random sampling but later than if it had been the only item with those features.

GS has gone through major revisions yielding GS2 (Wolfe, 1994) and GS3 (Wolfe & Gancarz, 1996). GS2 was an elaboration on GS1 seeking to explain new phenomena and to provide an account for the termination of search on target-absent trials. GS3 was an attempt to integrate the covert deployments of visual attention with overt deployments of the eyes. This paper describes the current state of the next revision, uncreatively dubbed Guided Search 4.0 (GS4). The model is not in its final state because several problems remain to be resolved.

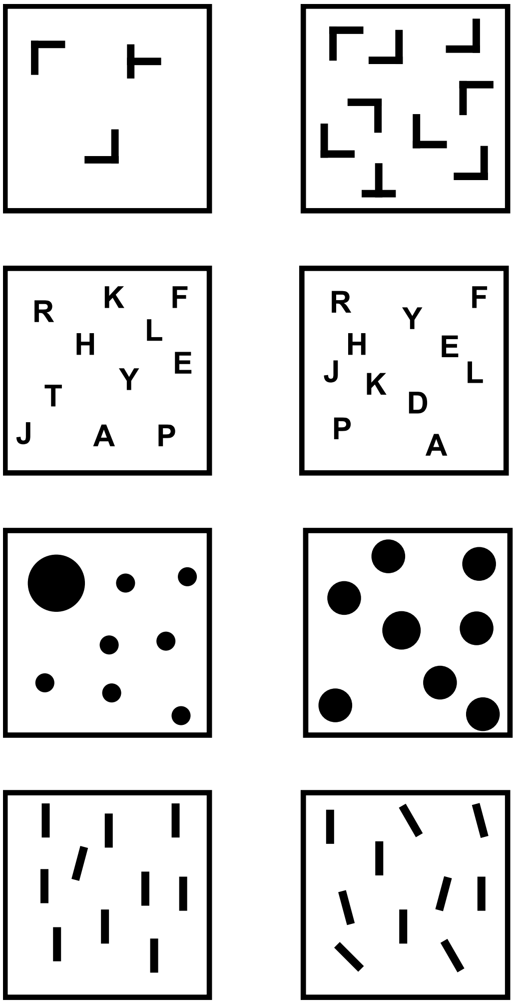
## What Does Guided Search 4.0 Seek to Explain?

GS4 is a model of simple search tasks done in the laboratory with the hope that the same principles will scale up to the natural and artificial search tasks that are performed continuously by people outside of the laboratory. A set of phenomena is described here. Each pair of figures illustrates an aspect of the data that any comprehensive model of visual search should strive to account for (see Figure 8.1). The left-hand member of the pair is the easier search in each case.

In addition, there are other aspects of the data, not illustrated here, that GS4 seeks to explain. For example, a good model of search should account for the distributions and not merely the means of reaction times and it should explain the patterns of errors (see, e.g., Wolfe, Horowitz, & Kenner, 2005).

## The Structure of GS4

Figure 8.2 shows the current large-scale architecture of the model. Referring to the numbers on the figure, parallel processes in early vision (1) provide input to object recognition processes (2) via a mandatory selective bottleneck (3). One object or, perhaps, a group of objects can be selected to pass through the bottleneck at one time. Access to the bottleneck is governed by visual selective attention. *Attention* covers a very wide range of processes in the nervous system (Chun & Wolfe, 2001; Egeth & Yantis, 1997; Luck & Vecera, 2002; Pashler, 1998a, 1998b; Styles, 1997). In this chapter, we will use the term *attention* to refer to the control of selection at this particular bottleneck in visual processing. This act of selection is mediated by a "guiding representation," abstracted from early vision outputs (4). A limited number of attributes (perhaps 1 or 2 dozen) can guide the deployment of attention. Some work better than others. Guiding attention on the basis of a salient color works very well. Search for a red car among blue and gray ones will not be hard (Green & Anderson, 1956; Smith, 1962). Other attributes, such as *opacity* have a weaker ability to guide attention (Mitsudo, 2002; Wolfe, Birnkrant, Horowitz, & Kunar, 2005). Still others, like the presence of an intersection, fail to guide altogether (Wolfe & DiMase, 2003). In earlier versions of GS, the output of the first, preattentive stage guided the second attentive stage. However, GS4 recognizes that guidance is a control




**A. Set size:** All else being equal, it will be harder and will take longer to find a target (a *T* in this example) among a greater number of distractors than lesser (Palmer, 1995).

**B. Presence/absence:** Under most circumstances, it will take longer on average to determine that targets (again *T*) are absent than to determine that they are present (Chun & Wolfe, 1996).

**C. Features and target–distractor similarity:** A limited set of basic attributes support very efficient search (Wolfe & Horowitz, 2004). The larger the difference between target (here, a large disk) and distractors, the more efficient the search (Duncan & Humphreys, 1989).

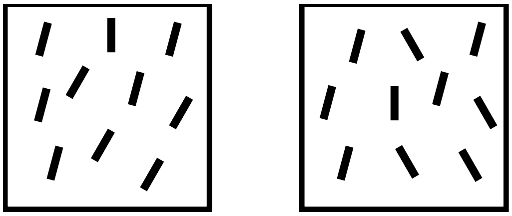
**D. Distractor heterogeneity:** The more heterogeneous the distractors, the harder the search (Duncan & Humphreys, 1989). Note that this is true in this example, even though the heterogeneous distractors are less similar to the target (line tilted to the right) than the homogeneous (Rosenholtz, 2001).

FIGURE 8.1 Eight phenomena that should be accounted for by a good model of visual search.

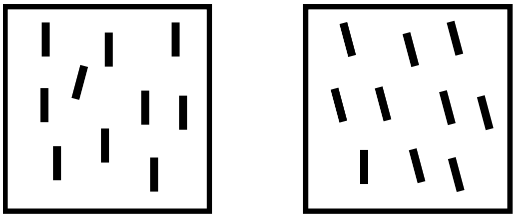
signal, derived from early visual processes. The guiding control signal is not the same as the output of early vision and, thus, is shown as a separate guiding representation in Figure 8.2 (Wolfe & Horowitz, 2004).

Some visual tasks are not limited by this selective bottleneck. These include analysis of image statistics (Ariely, 2001; Chong & Treisman, 2003) and some aspects of scene analysis (Oliva & Torralba, 2001). In Figure 8.2, this is shown as a second pathway, bypassing the selective bottleneck (5). It seems likely that selection can be guided by scene properties extracted in this second pathway (e.g., where are people likely to be in this image? [Oliva, Torralba, Castelhano, & Henderson, 2003]) (6). The notion that scene statistics can guide deployments of attention is a new feature of GS4. It is clearly related to the sorts of top-down or *reentrant* processing found in models like the Ahissar and Hochstein reverse hierarchy model (Ahissar & Hochstein, 1997; Hochstein & Ahissar, 2002) and the DiLollo et al. reentrant model (Di Lollo, Enns, & Rensink, 2000). These higher-level properties are acknowledged but not explicitly modeled in GS4.

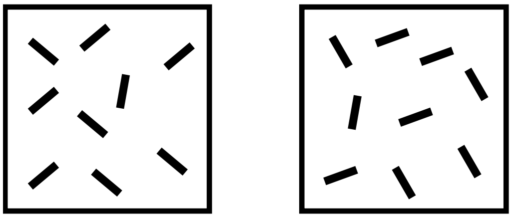
Outputs of both selective (2) and nonselective (5) pathways are subject to a second bottleneck (7). This is the bottleneck that limits performance in attentional blink (AB) tasks (Chun & Potter, 1995; Shapiro, 1994). This is a good moment to reiterate the idea that attention refers to several different processes, even in



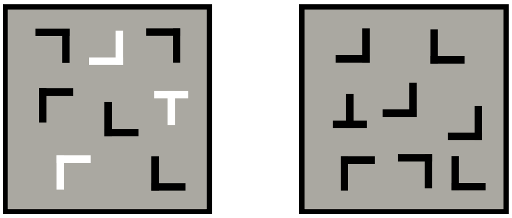
**E. Flanking/linear separability:** For the same target–distractor distances, search is harder when distractors flank the target. In this case, 0 deg among +15 and −30 is harder than 0 vs. +15 and +30. See linear separability in the two-dimensional color plane (Bauer, Jolicoeur, & Cowan, 1996).



**F. Search asymmetry:** Search for *A* among *B* is often different than search for *B* among *A*. Here 0 among −15 deg is harder than −15 among 0 (Rosenholtz, 2001; Treisman & Souther, 1985; Wolfe, 2001).



**G. Categorical processing:** All else being equal, targets are easier to find if they are categorically unique. On the left, the “steep” target is easier to find than the “steepest” target on the right. The geometric relationships are constant (Wolfe, Friedman-Hill, Stewart, & O’Connell, 1992).



**H. Guidance:** Of course, GS must explain guidance. It is easier to find a white *t* on the left than to find the *t* on the right. Color/polarity guides attention (Egeth, Virzi, & Garbart, 1984; Wolfe, Cave, & Franzel, 1989).

FIGURE 8.1 Continued.

the context of visual search. In AB experiments, directing attention to one item in a rapidly presented visual sequence can make it difficult or impossible to report on a second item occurring within 200–500 ms of the first. Evidence that AB is a late bottleneck comes from experiments that show substantial processing of “blinked” items. For example, words that are not reported because of AB can, nevertheless, produce semantic priming (Luck, Vogel, & Shapiro, 1996).

Object meaning does not appear to be available before the selective bottleneck (3) in visual search (Wolfe & Bennett, 1997), suggesting that the search bottleneck lies earlier in processing than the AB bottleneck (7). Moreover, depending on how one uses the term, *attention*, a third variety occurs even earlier in visual search. If an observer is looking for something red, all red items will get a boost that can be measured psychophysically (Melcher, Papathomas, & Vidnyánszky, 2005) and physiologically (Bichot, Rossi, & Desimone, 2005). Melcher et al. (2005) call this *implicit attentional selection*. We call it *guidance*. In either case, it is a global process, influencing many items at the same time—less a bottleneck than a filter. The selective bottleneck (3) is more local, being restricted to one object or location at a time (or, perhaps, more than one; McMains & Somers, 2004). Thus, even in the limited realm illustrated in Figure 8.2, attentional processes can be acting on early parallel stages (1) to select features, during search to select objects (3), and late, as part of decision or response mechanisms (7).

Returning to the selective pathway, in GS, object recognition (2) is modeled as a diffusion process where

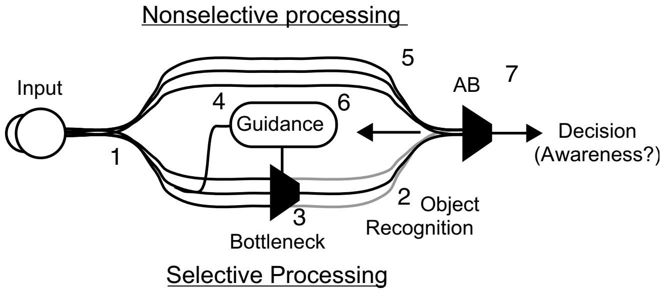



FIGURE 8.2 The large-scale structure of GS4. Numbers refer to details in text. Multiple lines illustrate parallel processing.

information accumulates over time (Ratcliff, 1978). A target is identified when information reaches a target threshold. Distractors are rejected when information reaches a distractor threshold. Important parameters include the rate and variability of information accrual and the relative values of the thresholds. Many parallel models of search show similarities to diffusion models (Dosher, Han, & Lu, 2004). Effects of set size on reaction time are assumed to occur either because accrual rate varies inversely with set size (limited-capacity models; Thornton, 2002; Figure 8.3) or because, to avoid errors, target, and distractor thresholds increase with set size (e.g., Palmer, 1994; Palmer & McLean, 1995).

In a typical parallel model, accumulation of information begins for all items at the same time. GS differs from these models because it assumes that information accumulation begins for each item only when it is selected (Figure 8.3). That is, GS has an *asynchronous* diffusion model at its heart. If each item needed to wait for the previous item to finish, this becomes a strict serial process. If *N* items can start at the same time, then this is a parallel model for set sizes of *N* or less. In its general form, this is a hybrid model with both serial and parallel properties. As can be seen in Figure 8.3, items are selected, one at a time, but multiple items can be accumulating information at the same time. A carwash is a useful metaphor. Cars enter one at a time, but several cars can be in the carwash at one time (Moore & Wolfe, 2001; Wolfe, 2003). (Though note that Figure 8.3 illustrates an unusual carwash where a car entering second could, in principle, finish first.)

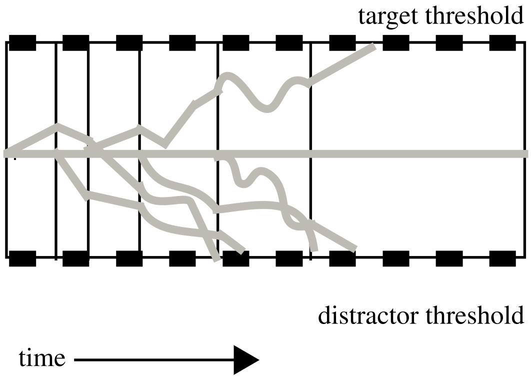



FIGURE 8.3 In GS4, the time course of selection and object recognition is modeled as an asynchronous diffusion process. Information about an item begins to accumulate only after that item has been selected into the diffuser.

As noted at the outset, search tasks have been modeled as either serial or parallel (or, in our hands, *guided*). It has proved difficult to use RT data to distinguish serial from parallel processes (Townsend, 1971, 1990; Townsend & Wenger, 2004). Purely theoretical considerations aside, it may be difficult to distinguish parallel from serial in visual search tasks because those tasks are, in fact, a combination of both sorts of process. That, in any case, is the claim of GS4, a model that could be described as a parallel–serial hybrid. It has a parallel front end, followed by an attentional bottleneck with a serial selection rule that then feeds into parallel object recognition processes.

### Modeling Guidance

In GS4, objects can be recognized only after they have been passed through the selective bottleneck between early visual processes and object recognition processes. Selection is controlled by a guiding representation. That final guiding representation is created bottom-up and top-down information. Guidance is not based directly on the contents of early visual processes but on a coarse and categorical representation derived from those processes. Why argue that guidance is a control process, sitting, as it were, to the side of the main selective pathway? The core argument is that information that is available in early vision (Figure 8.2, no. 1) and later (2) is not available to guidance (4). If guidance were a filter in the pathway, we would need to explain how information was lost and then regained (Wolfe & Horowitz, 2004).

Consider three examples that point toward this conclusion:

1. Even in simple feature search, efficient guidance requires fairly large differences between targets and distractors. For example, while we can resolve orientation differences on the order of a degree (Olzak & Thomas, 1986), it takes about a 15-deg difference to reliably attract attention (Foster & Ward, 1991b; Moraglia, 1989). Fine-grain orientation information is available before attentional selection and after but not available to the guidance mechanism.
2. Search is more efficient if a target is categorically unique. For example, it is easier to find a line that is the only "steep" item as illustrated in Figure 8.1. There is no categorical limitation on processing outside of the guidance mechanism.
3. Intersection type (*t*-junction vs. *x*-junction) does not appear to guide attention (Wolfe & DiMase, 2003). It can be used before selection to parse the field into preattentive objects (Rensink & Enns, 1995). Intersection type is certainly recognized in attentive vision, but it is not recognized by guidance.

Thus, we suggest that the guiding representation should be seen as a control module sitting to one side of the main selective pathway rather than as a stage within that pathway. In the current GS4 simulation, guidance is based on the output of a small number of broadly tuned channels. These can be considered to be channels for *steep*, *shallow*, left, and *right* (for steep and shallow, at least see Foster & Ward, 1991a). Only orientation and color are implemented, but other attributes are presumed to be similar. In orientation, the four channels are modeled as the positive portion of sinusoidal functions, centered at 0 (vertical), 90, 45, and −45 deg and raised to a power less than 1.0 to make the tuning less sharp. Thus, the steep channel is defined
[AQ1] as $\max[\cos(2^*\mathrm{deg}), 0]^{0.3}$. The precise shape is not critical for the qualitative performance of the model. In color, a similar set of channels covers a red-green axis with three categorical channels for *red*, *yellow*, and *green*. Color, of course, is a three-dimensional feature space. Restricting modeling to one red-green axis is merely a matter of convenience.

Another major simplification needs to be acknowledged. Selection is presumed to select objects (Wolfe & Bennett, 1997). As a consequence, the "receptive field" for the channels described above is an object, conveniently handed to the model. The model does not have a way to parse a continuous image into "preattentive object files" (our term) or "proto-objects" (Rensink & Enns, 1995, 1998).

### Bottom-Up Guidance

The more an item differs from its neighbors, the more attention it will attract, all else being equal. This can be seen in Figure 8.4. The vertical line "pops out" even though you were not instructed to look for vertical. That this pop-out is the result of local contrast can be intuited by noticing that the other four vertical lines in this image do not pop-out. They are not locally distinct (Nothdurft, 1991, 1992, 1993).

In GS4, bottom-up salience for a specific attribute such as orientation is based on the differences between the channel response for an item and the other items in the field. Specifically, for a given item, in orientation, we calculate the difference between the response to the item and the response to each other item for each of the four categorical channels. For each pairwise comparison, it is the maximum difference that contributes to bottom-up salience. The contribution of each pair is divided by the distance between the items. Thus, closer neighbors make a larger contribution to bottom-up

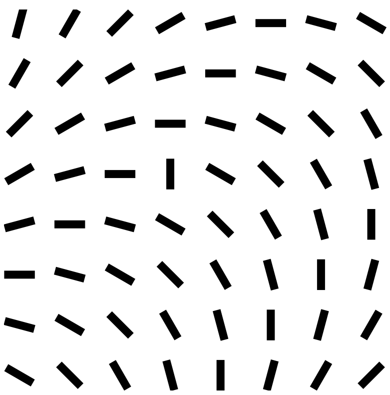


FIGURE 8.4 Local contrast produces bottom-up guidance. Note that there are five vertical lines in this display. Only one is salient.

activation of an item than do more distant items (Julesz, 1981, 1984). The distance function can be something other than linear distance. In the current simulation, we actually use the square root of the linear distance. Further data would be needed to strongly constrain this variable.

$$\sum_{b=1}^{\text{setsize}} \{\max[\mathrm{Ch}_1(a) - \mathrm{Ch}_1(b))\ldots (\mathrm{Ch}_n(a) - \mathrm{Ch}_n(b))]/\,d_{ab}\} \quad \text{Bottom-up activation}$$

Thus, this bottom-up calculation will create a bottom-up salience map where the signal at each item's location will be a function of that item's difference from all other items scaled by the distance between items.

Local differences are the basis for many models of stimulus salience (e.g., Itti & Koch, 2000; Koch & Ullman, 1985; Li, 2002). Many of these use models of cells in early stages of visual processing to generate signals. In principle, one of these salience models could replace or modify the less physiologically driven bottom-up guidance modules in GS4.

### Top-Down Guidance

If you were asked to find the targets in Figure 8.5, it would be reasonable to ask, "What targets?" However, if told to find the horizontal items, you can rapidly locate them. Thus, in Figure 8.5, bottom-up salience does not define targets, but efficient search is still possible, guided by top-down information. In GS4, top-down guidance is based on the match between a stimulus and the desired properties of the target. For each item, the channel responses are the signals out of which top-down guidance is created. The steep channel would respond strongly to the vertical lines, the "right" channel to 45-deg lines and so on. Top-down guidance results when higher weight is placed on the output of one channel than on others. In the current formulation of GS4, the model picks one channel for each attribute by asking which channel contains the largest signal favoring the target over the mean of the distractors. For example, consider a search for an orange line, tilted 22 deg off vertical. If the distractors were yellow and vertical, GS4 would place its weights on the red channel (targets and distractors both activate the yellow but only orange activates red) and the right-tilted channel (for similar reasons). If the same target were placed among red 45-deg lines, then it would be the yellow and steep channels that would contain the best signal.

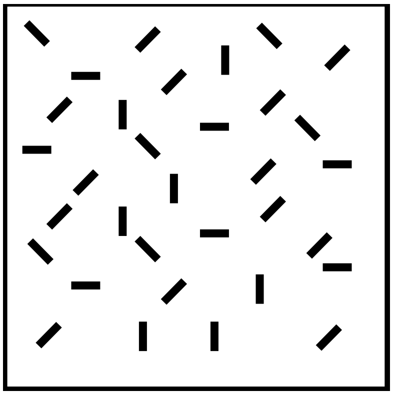


FIGURE 8.5 Bottom-up information does not define a target here, but top-down guidance can easily direct attention to a specified orientation (e.g., horizontal).

### The Activation Map

In GS, the activation map is the signal that will guide the deployment of attention. For each item in a display, the guiding activation is simply a weighted sum of the bottom-up activation and the activity in each channel (composed of the top-down activation) plus some noise. In the current version of GS, the weights are constrained so that one weight for a particular dimension (color or orientation) is set to 1.0 and the others are set to 0. This is the formal version of the claim that you can only select one feature in a dimension at a time (Wolfe et al., 1990). If you set the bottom-up weight to 0, you are making the claim that a salient but irrelevant distractor can be ignored. If you declare that it cannot go to 0, you are holding out the possibility of true *attentional capture* against the desires of the searcher. There is an extensive and inconclusive literature on this point (e.g., Bacon & Egeth, 1994; Folk, 1992; Lamy & Egeth, 2003; Theeuwes, 1994; Todd & Kramer, 1994; Yantis, 1998) that has been usefully reviewed by Rauschenberger (2003). GS4 does not allow the bottom-up weight to go to 0.

$$[\mathrm{wt}_{bu}(\mathrm{BU})] + \sum^{\text{Channels}} [\mathrm{wt}_{ch}(\mathrm{CH})] + \text{noise}$$

Activation Map

In earlier versions of GS, the activation map was fixed for a trial. Attention was deployed in order of activation strength from highest down until the target was found or until the search was abandoned. This assumes perfect memory for which items have been attended. Subsequent work has shown this to be incorrect (Horowitz & Wolfe, 1998, 2005). More will be said on this topic later. For now, the relevant change in GS4 is that the added noise is dynamic and each deployment of attention is directed to the item with the highest current activation.

## Guidance and Signal Detection Theory

Note that GS4, to this point, is very similar to a signal detection theory (SDT) model (Cameron, Tai, Eckstein, & Carrasco, 2004; Palmer & McLean, 1995; Palmer, Verghese, & Pavel, 2000; Verghese, 2001). Consider the standard SDT-style experiment. A search display is presented for 100 ms or so and masked. The distractors can be thought of as noise stimuli. The target, if present, is signal plus noise. In a standard SDT account, the question is how successfully the observer can distinguish the consequence of $N$(noise) from $[(N-1)$ (noise) + signal], where $N$ is the set size. As $N$ gets larger, this discrimination gets harder and that produces set size effects in brief exposure experiments. SDT models generally stop here, basing a decision directly on the output of this parallel stage. In GS, the output of this first stage guides access to the second stage. However, for brief stimulus presentation, GS4, like SDT models, would show a decrease in accuracy, albeit via a somewhat different mechanism. With a brief exposure, success in GS depends on getting attention to the target on the first deployment (or in the first few deployments). If there is no guiding signal, the chance of deploying to the target first is $1/N$. Performance drops as $N$ increases. As the guiding signal improves, the chance of deploying to the target improves. If the signal is very large, the effect of increasing $N$ becomes negligible and attention is deployed to the target first time, every time. There is more divergence between the models when stimulus durations are long. The rest of the GS model deals with deployments of attention over a more extended period. SDT models have not typically addressed this realm (but see Palmer, 1998). GS rules make different quantitative predictions than SDT "max" or "sum" rules but these have not been tested as yet.

### Why Propose a Bottleneck?

GS is a two-stage model with the activation map existing for the purpose of guiding access to the second stage where object recognition occurs. Why have two stages? Why not base response on a signal derived, like the activation map, in parallel from early visual processes? Single-stage models of this sort account for much search performance, especially for briefly presented stimuli (Baldassi & Burr, 2000; Baldassi & Verghese, 2002; McElree & Carrasco, 1999; Palmer & McLean, 1995). Is there a reason to propose a bottleneck in processing with access controlled by guidance? Here are four lines of argument, which, taken together, point to a two-stage architecture.

1. *Targets may be easy to identify but hard to find*. Consider the search for a *T* among *Ls* in Figure 8.1A and the search for tilted among vertical in Figure 8.1D. In isolation, a *T* is trivially discriminable from an *L* and tilted is trivially discriminable from vertical. However, search for the *T* is inefficient while search for tilted is efficient. The GS, two-stage account is fairly straightforward. The first stage registers the same vertical and horizontal elements for *Ts* and *Ls*. However, the intersection type is not available to guide attention (Wolfe & DiMase, 2003). The best that guidance can do is to deliver one object after another to the second stage. The relationship between the vertical and horizontal elements that identifies an object as *T* or *L* requires second-stage binding. The lack of guidance makes the search inefficient. The orientation search in 1d, in contrast, is easy because the first stage can guide the second stage. This argument would be more convincing if the single T and the tilted line were equated for discriminability. Even so, a single stage model must explain why one easy discrimination supports efficient search and another does not.

2. *Eye movements.* Saccadic eye movements impose an obvious seriality on visual processing (Sanders & Houtmans, 1985). Attention is deployed to the locus of the next saccade before it is made (Kowler, Anderson, Dosher, & Blaser, 1995), and guidance mechanisms influence the selection of eye movement targets (Bichot & Schall, 1999; Motter & Belky, 1998; Shen, Reingold, & Pomplun, 2003; Thompson & Bichot, 2004).

Invoking the control of saccades as an argument for a model of covert deployments of attention is a double-edged sword. Numerous researchers have argued that overt deployment of the eyes is what needs to be explained and that there is no need for a separate notion of covert deployments (Deubel & Schneider, 1996; Findlay & Gilchrist, 1998; Maioli, Benaglio, Siri, Sosta, & Cappa, 2001; Zelinsky & Sheinberg, 1995, 1996). If true, the link between attention and eye movements is not trivially simple. Take the rate of processing for example. The eyes can fixate on 4–5 items per second. Estimates of the rate of processing in visual search are in the range of 10 to 30 or 40 per second (based, e.g., on search slopes). The discrepancy can be explained by assuming that multiple items are processed, in parallel, on each fixation. Indeed, it can be argued that eye movements are a way for a parallel processor to optimize its input, given an inhomogeneous retina (Najemnik & Geisler, in press).

Eye movements are not required for visual search. With acuity factors controlled, RTs are comparable with and without eye movements (Klein & Farrell, 1989; Zelinsky & Sheinberg, 1997), and there is endless evidence from cueing paradigms that spatial attention can be deployed away from the point of fixation (for useful reviews, see Driver, 2001; Luck & Vecera, 2002). Nevertheless, the neural circuitry for eye movements and for deployment of attention are closely linked (Moore, Armstrong, & Fallah, 2003; Schall & Thompson, 1999), so the essential seriality of eye movements can point toward the need for a serial selection stage in guided search.

3. *Binding*. The starting point for Treisman's feature integration theory was the idea that attention was needed to *bind* features together (Treisman & Gelade, 1980). Failure to bind correctly could lead to *illusory conjunctions*, in which, for example, the color of one object might be perceived with the shape of another (Treisman & Schmidt, 1982). While the need for correct binding can be seen as a reason for restricting some processing to one item at a time, it is possible that multiple objects could be bound at the same time. Wang, for example, proposes an account where correlated oscillations of activity are the mechanism for binding and where several oscillations can coexist (Wang, 1999) and Hummel & Stankiewicz (1998) showed that a single parameter that varies the amount of overlap between oscillatory firings acts a lot like attention. The oscillation approach requires that when several oscillations coexist, they must be out of synchrony with each other to prevent errors like illusory conjunctions. Given some required temporal separation between oscillating representations, this places limit on the number of items that can be processed at once, consistent with an attentional bottleneck.

4. *Change blindness*. In change blindness experiments, two versions of a scene or search display alternate. If low-level transients are hidden, observers are poor at detecting substantial changes as long as those changes do not alter the gist, or meaning, of the display (Rensink, O'Regan, & Clark, 1997; Simons & Levin, 1997; Simons & Rensink, 2005). One way to understand this is to propose that observers only recognize changes in objects that are attended over the change and that the number of objects that can be attended at one time is very small, perhaps only one. In a very simple version of such an experiment, we asked observers to examine a display of red and green dots. On each trial, one dot would change luminance. The Os' task was to determine whether it also changed color at that instant. With 20 dots on the screen, performance was 55% correct. This is significantly above the 50% chance level but not much. It is consistent with an ability to monitor the color of just 1–3 items (Wolfe, Reinecke, & Brawn, 2006).

Early vision is a massively parallel process. So is object recognition. A stimulus (e.g., a face) needs to be compared with a large set of stored representations in the hopes of a match. The claim of two-stage models is that there are profound limitations on the transfer of information from one massively parallel stage to the next. Those limitations can be seen in phenomena such as change blindness. At most, it appears that a small number of objects can pass through this bottleneck at one time. It is possible that the limit is one. Guidance exists to mitigate the effects of this limitation. Under most real-world conditions, guidance allows the selection of an intelligently chosen subset of all possible objects in the scene.

## Modeling the Bottleneck

In earlier versions of GS, object recognition was regarded as something that happened essentially instantaneously when an item was selected. That was never intended to be realistic. Data accumulating from many labs since that time has made it clear that the time required to identify and respond to a target is an important constraint on models of the bottleneck in the selective pathway. If it is not instantaneous, how long is selective attention occupied with an item after that item is selected? Measures of the *attentional dwell time* (Moray, 1969) have led to apparently contradictory results. One set of measures comes from attentional blink (Raymond, Shapiro, & Arnell, 1992; Shapiro, 1994) and related studies (Duncan, Ward, & Shapiro, 1994; Ward, Duncan, & Shapiro, 1996, 1997). These experiments suggest that, once attention is committed to an object, it is tied up for 200–500 ms (see also Theeuwes, Godijn, & Pratt, 2004). This dwell time is roughly consistent with the time required to make voluntary eye movements and volitional deployments of attention (Wolfe, Alvarez, & Horowitz, 2000). It would seem to be incompatible with estimates derived from visual search. In a classic, serial self-terminating model of search, the time per item is given by the slope of target-absent trials or twice the slope of the target-present trials. Typical estimates are in the range of 30–60 ms/item. Efforts have been made to find a compromise position (Moore, Egeth, Berglan, & Luck, 1996), but the real solution is to realize that slopes of RT × set size functions are measures of the rate of processing, not of the time per item. We have made this point using a carwash metaphor (Moore & Wolfe, 2001; Wolfe, 2002; cf. Murdock, Hockley, & Muter, 1977). The core observation is that, while cars might enter (or emerge from) a carwash at a rate of 50 ms/item, they might be in this very fast carwash for 200–500 ms. Of course, a necessary corollary of this observation is that more than one car can be in the carwash at one time.

In GS4, as noted earlier, the carwash is formally modeled with an asynchronous diffusion model. Asynchronous diffusion is really a class of models with a large number of parameters, as illustrated in Figure 8.6. Having many parameters is not usually seen as a strength of a model (Eckstein, Beutter, Bartroff, & Stone, 1999). However, complex behaviors are likely to have complex underpinnings. The goal of this modeling effort is to constrain the values of the parameters

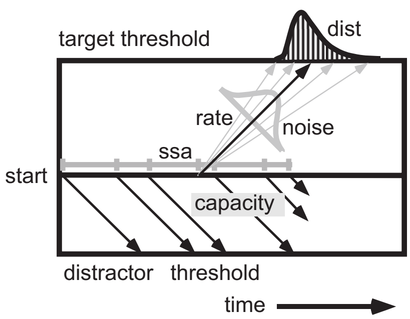



FIGURE 8.6 The parameters of an asynchronous diffusion model.

so that variation in a small subset can account for a large body of data.

The assumption of diffusion models is that information begins to accumulate when an item is selected into the diffuser. The time between successive selections is labeled *ssa* for *stimulus selection asynchrony*. It could be fixed or variable. In either case, the average SSA is inversely related to the rate of processing that, in turn, is reflected in the slope of RT × set size functions. Because search RT distributions are well described as gamma distributions, we have used exponentially distributed interselection intervals. However, it is unclear that this produces a better fit to the data than a simple, fixed interval of 20–40 ms/item.

In the case of visual search, the goal is to determine if the item is a target or a distractor and the answer is established when the accumulating information crosses a *target threshold* or *distractor threshold*. Both of those thresholds need to be set. It would be possible to have either or both thresholds change over time (e.g., one might require less evidence to reject a distractor as time progresses within a search trial). In the present version of GS4, the target threshold, for reasons described later, is about 10 times the distractor threshold. The *start* point for accumulation might be fixed or variable to reflect a priori assumptions about a specific item. For example, contextual cueing effects might be modeled by assuming that items in the cued location start at a point closer to the target threshold (Chun, 2000; Chun & Jiang, 1998). In current GS4, the start point is fixed.

Items diffuse toward a boundary at some average *rate*. In principle, that rate could differ for different items in a display (e.g., as a function of eccentricity Carrasco, Evert, Chang, & Katz, 1995; Carrasco & Yeshurun, 1998; Wolfe, O'Neill, & Bennett, 1997). The rate divided into the distance to the threshold gives the average time in the diffuser for a target or distractor. The diffusion process is a continuous version of a random walk model with each step equal to the rate plus some *noise*. In current GS4, the rate parameter is used to account for differences between Os, but is set so that the time for a target to diffuse to the target boundary is on the order of 150–300 ms. Ratcliff has pointed out that noise that is normally distributed around the average path will produce a positively skewed *distribution* of finishing times (Ratcliff, 1978; Ratcliff, Gomez, & McKoon, 2004). This is a useful property since search RTs are positively skewed. An asynchronous diffusion model assumes that information about items can start accumulating at different times.

The diffuser is assumed to have some *capacity*. This brings with it a set of other choices that need to be made. If the capacity is $K$, then the $K + 1$th item cannot be selected until the one of the $K$ items is dismissed. At the start of a search, can $K$ items be selected simultaneously into an empty diffuser? If items are selected one at a time, then there will be periods when the number of items in the diffuser is less than $K$. This will also occur, of course, if the set size is less than $K$. When the diffuser contains fewer than $K$ items, is the rate of information accumulation fixed or is it proportional to the number of items in the diffuser. That is, if $K = 4$ and the set size is 2, does the processing rate double? In GS4, we typically use a capacity of 4 items (inspired, in part, by the ubiquity of the number 4 in such capacity estimates (Cowan, 2001). Small changes in $N$ do not produce large changes in the behavior of the model. At present, in GS4, if there are fewer than the maximum number of items in the diffuser or if the same item is selected more than once (hard for cars in a car wash but plausible here), then the rate of information accrual increases.

## Memory in Search

If capacity, $N$, is less than the set size, then the question of memory in search arises. If an item has been dismissed from the diffuser, can it be reselected in the same search? The classic serial, self-terminating model (FIT and earlier versions of GS) had a capacity of one (i.e., items are processed in series) and an assumption that items were not reselected. That is, visual search was assumed to be sampling without replacement. In 1998, we came to the conclusion that visual search was actually sampling with replacement—that there was no restriction on reselection of items (Horowitz & Wolfe, 1998). Others have argued that our claim that "visual search has no memory" was too strong and that selection of some number of recently attended items is inhibited (Kristjansson, 2000; Peterson, Kramer, Wang, Irwin, & McCarley, 2001; Shore & Klein, 2000). In our work, we have been unable to find evidence for memory in search. Nevertheless, we have adopted a middle position in our modeling. Following Arani, Karwan, and Drury (1984), the current version of GS inhibits each distractor as it is rejected. At every cycle of the model thereafter, there is some probability that the inhibition will be lifted. Varying that probability changes the average number of items that are inhibited. If that parameter is 1, then visual search has no memory. If it is 0, search has perfect memory. We typically use a value of 0.75. This yields an average of about three inhibited items at a time during a search trial. These are not necessarily the last three rejected distractors. Rigid $N$-back models of memory in search tend to make strong predictions that are easily falsified (e.g., that search through set sizes smaller than $N$ will show perfect memory). Modest variation in this parameter does not appear to make a large difference in model output.

## Constraining Parameter Values

At this point, the reader would be forgiven for declaring that a model with this many parameters will fit all possible data and that some other model with fewer parameters must be preferable. If all of the parameters could vary at will, that would be a fair complaint. However, GS assumes that most of these are fixed in nature; we just do not know the values. Moreover, other apparently simple models are simple either by virtue of making simplifying assumptions about these (or equivalent) parameters or by restriction of the stimulus conditions. For example, if stimuli are presented briefly, then many of the issues (and parameters) raised by an asynchronous diffusion process become moot.

The data provide many constraints on models of search. At present, it must be said, that these constraints are better at ruling out possibilities than they are at firmly setting parameters, but modeling by exclusion is still progress. We have obtained several large data sets in an effort to understand normal search behavior. Figure 8.7 shows average RTs for 10 Os, tested for 4,000 trials on each of three search tasks: A simple feature search for a red item among green distractors, a color X orientation conjunction search, and a "spatial configuration" search for a 2 among 5s, the mirror reverse of the 2. The 2 versus 5 search might have been called a *serial* search in the past but that implies a theoretical position. Calling it an inefficient spatial configuration search is neutrally descriptive. This data set, confirming other work (Wolfe, 1998), shows that the ratio of target-absent to target-present slopes is greater than 2:1 for spatial configuration searches. This violates the assumptions of a simple serial, self-terminating search model with complete memory for rejected distractors. The variance of the RTs increases with set size and is greater for target-absent

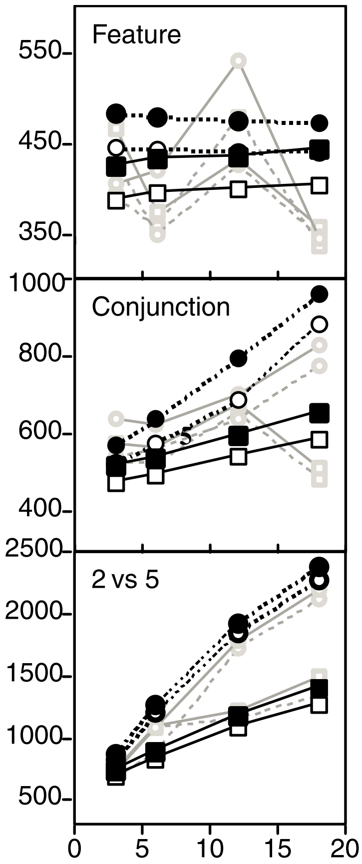



FIGURE 8.7 Average reaction times for 10 observers tested for 1,000 trials per set size in three tasks: Feature (red among green), conjunction (red vertical among red horizontal and green vertical), and a search for a 2 among 5 (the mirror-reversed item). The black, bold lines represent correct responses. Light-gray lines are the corresponding error trials. Squares are hits, circles are correct absent responses; closed symbols are means, open are medians (always slightly faster than the means). In the gray error trials, squares are false alarms (very rare), circles are misses. Note the very different *y*-axes.

than for target-present trials. Error rates increase with set size in all conditions (Figure 8.8). The great bulk of errors are miss errors: False alarms are rare in RT search studies. Miss-error RTs tend to be somewhat faster than correct absent RTs (Figure 8.7). Thus, if a model predicts a large number of false alarms or predicts that errors are slow, it is failing to capture the shape of the data. GS4 produces qualitatively correct patterns of RTs as described later.

In a separate set of experiments, we tested the same three tasks on a wider and denser range of set sizes than is typical. As shown in Figure 8.9, the salient finding is that RT × set size functions are not linear (Wolfe, Michod, & Horowitz, 2004). They appear to be compressive with small set sizes (1–4) producing very steep slopes. The cause of the nonlinearity is not clear but

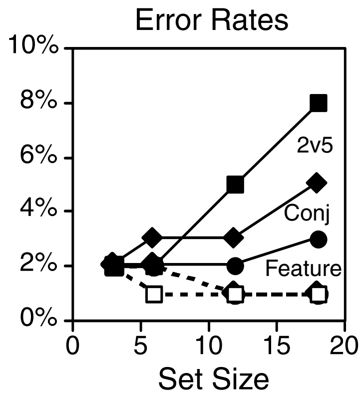



FIGURE 8.8 Average error rates for data shown in Figure 8.7. Closed symbols are miss errors as a percentage of all target-present trials. Open symbols are false alarms (all false alarm rates are low and similar).

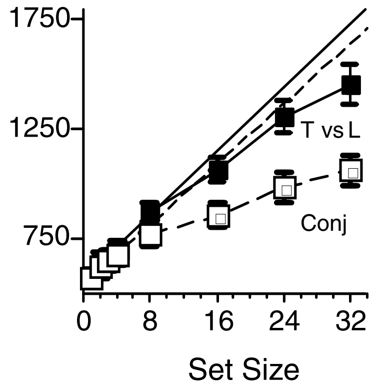



FIGURE 8.9 RT (reaction time) × set size functions with linear regression lines fitted to just Set Sizes 1–4 to illustrate the nonlinearity of these functions.

[AQ2]

the result means that models (like earlier versions of GS) that produce linear RT × set size functions are missing something. In GS4, a nonlinearity is produced by allowing the rate of information accrual to be proportional to the number of items in the diffuser (up to the capacity limit). Small set sizes will benefit more from this feature than large, causing small set size RTs to be somewhat faster than they would otherwise be.

The large number of trials that we ran to collect the data in Figures 8.7 and 8.8 allows us to look at RT distributions. Search RT distributions, like so many other RT distributions, are positively skewed (Luce, 1986; Van Zandt, 2002). This general shape falls out of diffusion models (Ratcliff et al., 2004). In an effort to compare distributions across Os, set sizes, and search tasks, we normalized the distributions using a nonparametric equivalent of a *z*-transform. Specifically, the 25th and 75th percentiles of the data were transformed to $-1$ and $+1$, respectively, and the data were scaled relative to the interquartile distance. As shown in Figure 8.10, the first striking result of this analysis is how similar the distribution shapes are. To a first approximation, distributions for feature and conjunction searches are scaled copies of each other with no qualitative change

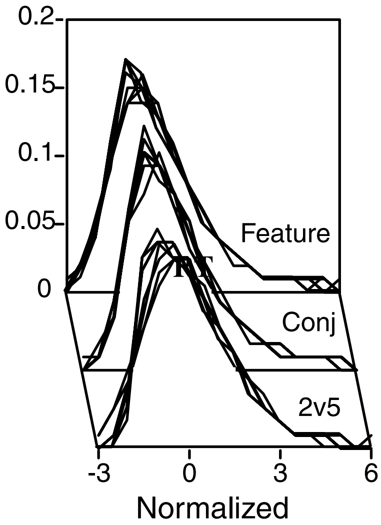



FIGURE 8.10 Probability density functions for normalized RT distributions for four set sizes in three search tasks. Thicker lines are target-present; thinner are target-absent. Note the similarity of the probability density functions, especially for the feature and conjunction tasks.

in the shape of the distribution with set size. Models that predict that the shape of the normalized RT distribution changes with set size would, therefore, be incorrect. Moreover, after this normalization, there is little or no difference between target-present (thick lines) and target-absent (thin-lines)—also a surprise for many models (e.g., FIT and earlier version of GS).

RT distributions from the 2 versus 5 task are somewhat different. They are a bit more rounded than the feature and conjunction distributions. They change a little with set size and absent distributions are somewhat different from present distributions. A number of theoretical distributions (gamma, Weibull, lognormal, etc.) fit the distributions well, and there does not seem to be a data-driven reason to choose between these at the present time. GS4 produces RT distributions that are qualitatively consistent with the pattern of Figure 8.10.

Hazard functions appear to magnify these differences. Hazard functions give the probability of finding the target at one time given that it has not been found up until that time. In Figure 8.11, we see that the hazard functions are clearly nonmonotonic. All tasks at all set sizes, target present or absent, seem to have the same initial rise. (The dashed line is the same in all three panels.) The tasks differ in the later portions of the curve, but note that data beyond an *x*-value of 3 come from the few trials in the long tail of this RT distribution. Gamma and ex-Gaussian distributions have monotonic hazard functions and, thus, are imperfect models of these RT distributions. Van Zandt and Ratcliff (2005) note that "the increasing then decreasing hazard is ubiquitous" and are an indication that the RT distribution is a mixture of two or more underlying distributions. This seems entirely plausible in the case of a complex behavior like search.

From the point of view of constraining models of search, a model should not predict qualitatively different shapes of RT distributions, after normalization, as a function of set size, task, or target presence or absence for reasonably efficient searches. Some differences between more efficient (feature and conjunction) and less efficient (2 vs. 5) search are justified. Moreover, inefficient search may produce some differences in distributions as a function of set size and presence/absence.

## Target-Absent Trials and Errors

In some ways, modeling the process that observers use to find a target is the easy part of creating a model of

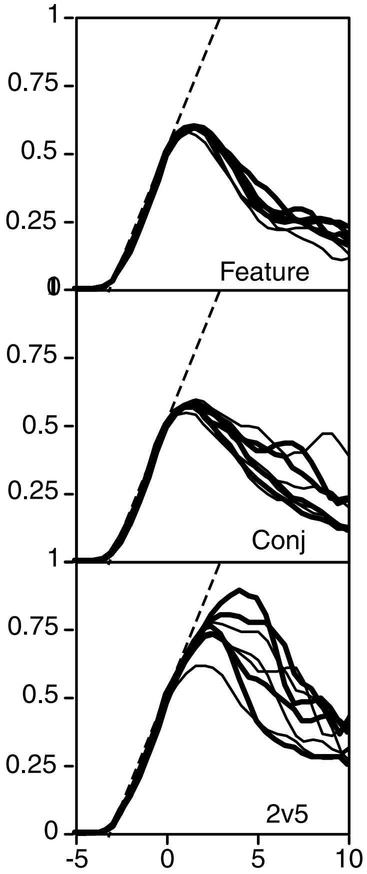



FIGURE 8.11 Hazard functions for the probability density functions in Figure 8.10.

visual search. After all, once attention has been guided to the target, the model's work is done. What happens when no target is present? When do you terminate an unsuccessful search? Simple serial models have a clear account. When you have searched all items, you quit. Such models predict lower variance on target absent trials than on target present trials because target absent trials should always require observers to attend to $N$ items where $N$ is the set size. On target present trials, observers might find a target on the first deployment of attention or on the $N$th. That prediction is not correct. Moreover, we had observers search through displays in which items were continuously being replotted in random locations and found that observers can terminate search under these conditions even though dynamic search displays would make it impossible to know when everything had been examined (Horowitz & Wolfe, 1998). (Note, compared with standard search tasks, dynamic search conditions do lead to longer target-absent RTs and more errors, suggesting some disruption of target-absent search termination.)

Instead of having a method of exhaustively searching displays, observers appear to establish a quitting rule in an adaptive manner based on their experience with a search task. Observers speed subsequent responses after correct responses and slow subsequent responses after errors (Chun & Wolfe, 1996). An adaptive rule of this sort can be implemented in many ways. Observers could adjust the time spent searching per trial. They could adjust the number of items selected or the number of items rejected. Whatever is adjusted, the resulting quitting threshold must be scaled by set size. That is, the threshold might specify quitting if no target has been found after some percentage of the total set size has been selected, not after some fixed number of items had been selected regardless of set size.

In GS4, miss errors occur when the quitting threshold is reached before the target is found. As shown in Figure 8.7, miss RTs are slightly faster than RTs for correct absent trials. Misses occur when observers quit too soon. As shown in Figure 8.8, false alarms are rare and must be produced by another mechanism. If observers produced false alarms by occasionally guessing "yes" when the quitting threshold was reached, then false alarms and miss RTs should be similar, which they are not. False alarms could be produced when information about distractor items incorrectly accumulates to the target boundary. There may also be some sporadic fast guesses that produce false alarms. At present, GS4 does not produce false alarms at even the infrequent rate that they are seen in the data.

The data impose a number of constraints on models of search termination. Errors increase with set size, at least for harder search tasks. One might imagine that this is a context effect. The quitting threshold gets set to the average set size and is, therefore, conservative for smaller set sizes and liberal for larger. This cannot be the correct answer because the patterns of RTs and errors do not change in any qualitative way when set sizes are run in blocks rather than intermixed (Wolfe, Palmer, Horowitz, & Michod, 2004). Slopes for target-absent are reliably more than twice as steep as slopes for target present trials (Wolfe, 1998).

One of the most interesting constraints on search termination is that observers appear to successfully terminate target-absent trials too fast. Suppose that observers terminated trials at time $T$, when they were convinced

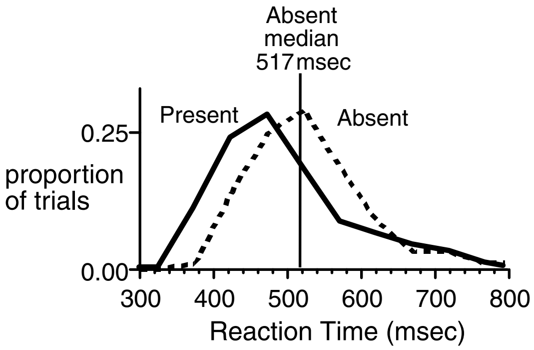



FIGURE 8.12 Reaction time (RT) distributions for one observer, Set Size 3, conjunction task. Note the degree of overlap between target-present and target-absent RTs. Twenty-five percent of correct present trials lie above the median for the correct absent trials. Miss error rate in this condition is 1.9%. How can observers answer "no" so quickly?

that only *X*% of targets would require more than *T* ms to find, where *X*% is the error rate (for the condition illustrated in Figure 8.12, the miss error rate is approximately 1.9%). While the details depend on the particulars of the model (e.g., assumptions about guessing rules and RT distributions), the median of the target absent RTs should cut off about *X*% of the target present distribution. A glance at Figure 8.12 shows that this is not true for one Os conjunction data for Set Size 3. More than 25% of the correct target present RTs lie above absent median. This is merely an illustrative example of a general feature of the data. The mean/median of the absent RTs falls far too early. This is especially true for the smaller set sizes where 30% of target-present RTs can fall above the target-absent mean. There are a variety of ways to handle this. Returning to Figure 8.6, it is reasonable to assume that the target threshold will be much higher than the distractor threshold. A Bayesian way to think about this is that an item is much more likely to be a distractor than a target in a visual search experiment. It is therefore reasonable to dismiss it as a distractor more readily than to accept it as a target. If observers can successfully quit after *N* distractors have been rejected, it is possible that a fast target-absent search could end in less time than a slow target-present search. The present version of GS uses this difference in thresholds to capture this aspect of the data. The ratio of target to distractor threshold is generally set to 10:1. Nevertheless, while we can identify these constraints in the data, we are still missing something in our understanding of blank trial search termination. Modeling the pattern of errors is the least successful aspect of GS4 at the present time. Parameters that work in one condition tend to fail in others.

## State of the Model

To what extent does GS4 capture the diverse empirical phenomena of visual search? Figure 8.13 shows data for the 2 versus 5 task for a real *O* (solid symbols) and for the model using parameters as described above (same diffusion parameters, same error rules, etc.). The free parameter is a rate parameter that is used to equate target present slopes so the excellent match between model

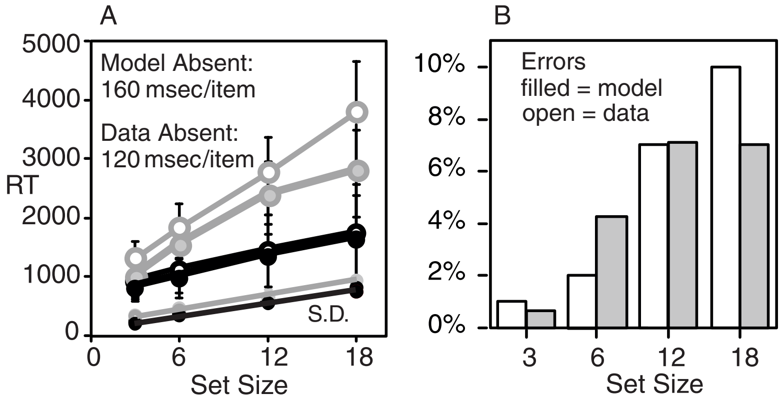



FIGURE 8.13 (A) An example of GS4 model data compared with one *O*'s data for the 2 vs. 5 task. Solid symbols indicate the *O*, open symbols the model. Target-present trials are in black, target-absent in gray. Small symbols denote standard deviations. (B) Miss error rates: open bars are data; filled are model results.

and data for the target present data is uninteresting. Target-absent RTs produced by the model are a reasonable approximation of the data, though the slopes are too steep. Standard deviations of the RTs (shown at the bottom of the figure) are very close for data and model. The model and the observer had very similar errors rates, rising from about 2% to about 8% as a function of set size. Model RT distributions are positively skewed and qualitatively similar to the real data.

If we now use exactly the same parameters for the conjunction tasks, the model produces slopes of 12 ms/item on target-present and 24 ms/item for target-absent trials. This compares well to 9 and 26, respectively for this Os data. However, the model's target-absent RTs are significantly too fast. Moving the distractor threshold is one way to compensate, but this disrupts slopes and errors. The model does not quite capture the Os rules for search termination. The heart of the problem seems to relate to the point illustrated by Figure 8.12. Real observers are somehow able to abandon unsuccessful searches quickly without increasing their error rates unacceptably. We have not developed a mechanism that allows GS4 to avoid this speed-accuracy tradeoff.

GS4 does capture other qualitative aspects of the search data, however. Returning to the checklist in Figure 8.1, the model certainly produces appropriate set size effects (Figure 8.1A) and differences between target-present and target-absent trials (Figure 8.1B). The structure of the first, guiding stage produces most of the other properties listed here. Search becomes less efficient as target-distractor similarity increases (Figure 8.1C) and as distractor heterogeneity increases (Figure 8.1D). If the target is flanked by distractors, the setting of top-down weights is less successful and efficiency declines (Figure 8.1E). If the target is defined by the presence of a categorical attribute, search is more efficient than if it is defined by the absence of that attribute (Figure 8.1G). Thus, for example, in search for 15 deg among 0 deg, GS4 can place its weight on the right-tilted channel and find a signal that is present in the target and absent in the distractors. If the target is 0 deg and the distractors are 15 deg, the best that can be done is to put weight on the "steep" channel. The 0-deg signal is bigger than the 15-deg signal in that channel but not dramatically. As a result, search is less efficient—a search asymmetry (Figure 8.1F). And, of course, guidance (Figure 8.1H) is the model's starting point. If the target is red, search will be guided toward red items.

## Summary

The current implementation of GS4 captures a wide range of search behaviors. It could be scaled up to capture more. The front end is currently limited to orientation and color (and only the red-green axis of color, at that). Other attributes could be added. This would allow us to capture findings about triple conjunctions, for example (Wolfe et al., 1989). Ideally, one of the more realistic models of early vision could be adapted to provide the front end for GS. At present, the guiding activation map is a weighted sum of the various sources of top-down and bottom-up guidance. The weights are set at the start of a block of trials. This is a bit simple-minded. A more complete GS model would learn its weights and would change them in response to changes in the search task. A more adaptive rule for setting weights could capture many of the priming effects in search. Observers would be faster to find a target if the target was repeated because the weights would have been set more effectively for that target (Hillstrom, 2000; Maljkovic & Nakayama, 1994; Wolfe, Butcher, Lee, & Hyle, 2003; though others might differ with this account; Huang, Holcombe, & Pashler, 2004). A more substantive challenge is presented by the evidence that attention is directed toward objects. While it would not be hard to imagine a GS front-end that expanded the list of guiding attributes beyond a cartoon of color and orientation processing, it is hard to envision a front end that would successfully parse a continuous image into its constituent *objects of attention*. The output of such front-end processing could be fed through a GS-style bottleneck to an object recognition algorithm. Such a model might be able to find what it was looking for but awaits significant progress in other areas of vision research. In the meantime, we believe that the GS architecture continues to serve as a useful model of the bottleneck between visual input and object recognition.

## REFERENCES

Ahissar, M., & Hochstein, S. (1997). Task difficulty and visual hierarchy: Counter-streams in sensory processing and perceptual learning. *Nature*, 387(22 May 1997), 401–406.

Arani, T., Karwan, M. H., & Drury, C. G. (1984). A variable-memory model of visual search. *Human Factors*, 26(6), 631–639.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception and Psychophysics*, *55*(5), 485–496.

Baldassi, S., & Burr, D. C. (2000). Feature-based integration of orientation signals in visual search. *Vision Research*, *40*(10–12), 1293–1300.

———, & Verghese, P. (2002). Comparing integration rules in visual search. *Journal of Vision*, *2*(8), 559–570.

Bauer, B., Jolicœur, P., & Cowan, W. B. (1996). Visual search for colour targets that are or are not linearly-separable from distractors. *Vision Research*, *36*(10), 1439–1466.

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*(5721), 529–534.

———, & Schall, J. D. (1999). Saccade target selection in macaque during feature and conjunction. *Visual Neuroscience*, *16*, 81–89.

Cameron, E. L., Tai, J. C., Eckstein, M. P., & Carrasco, M. (2004). Signal detection theory applied to three visual search tasks—identification, yes/no detection and localization. *Spatial Vision*, *17*(4–5), 295–325.

Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception and Psychophysics*, *57*(8), 1241–1261.

———, & Yeshurun, Y. (1998). The contribution of covert attention to the set size and eccentricity effects in visual search. *Journal Experimental Psychology: Human Perception and Performance*, *24*(2), 673–692.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, *22*, 225–271.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

Chun, M., & Jiang, Y. (1998). Contextual cuing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71.

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*, 170–178.

———, & Potter, M. C. (1995). A two-stage model for multiple target detection in RSVP. *Journal of Experimental Psychology: Human Perception & Performance*, *21*(1), 109–127.

———, & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*, 39–78.

———. (2001). Visual Attention. In E. B. Goldstein (Ed.), Blackwell's handbook of perception (pp. 272–310). Oxford: Blackwell.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114; discussion, 114–185.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual pathways. *Journal of Experimental Psychology: General*, *129*(3), 481–507.

Dosher, B. A., Han, S., & Lu, Z. L. (2004). Parallel processing in visual search asymmetry. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 3–27.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, 92, 53–78.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.

———, Ward, R., & Shapiro, K. (1994). Direct measurement of attention dwell time in human vision. *Nature*, *369*(26 May), 313–314.

Eckstein, M., Beutter, B., Bartroff, L., & Stone, L. (1999). Guided search vs. signal detection theory in target localization tasks. *Investigative Ophthalmology & Visual Science*, *40*(4), S346.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 32–39.

———, & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology*, *48*, 269–297.

Egeth, H. W. (1966). Parallel versus serial processes in multidimentional stimulus discrimination. *Perception and Psychophysics*, *1*, 245–252.

Findlay, J. M., & Gilchrist, I. D. (1998). Eye guidance and visual search. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 295–312). Amsterdam: Elsevier.

Folk, C. L., Remington, R. W., & Johnston J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044.

Foster, D. H., & Ward, P. A. (1991a). Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society (London B)*, *243*, 75–81.

——, & Ward, P. A. (1991b). Horizontal-vertical filters in early vision predict anomalous line-orientation frequencies. *Proceedings of the Royal Society (London B)*, *243*, 83–86.

Green, B. F., & Anderson, L. K. (1956). Color coding in a visual search task. *Journal of Experimental Psychology*, *51*(1), 19–24.

Hillstrom, A. P. (2000). Repetition effects in visual search. *Perception and Psychophysics*, *62*(4), 800–817.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*, 791–804.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*(August 6), 575–577.

——, & Wolfe, J. M. (2005). Visual search: The role of memory for rejected distractors. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 264–268). San Diego, CA: Academic Press/Elsevier.

Huang, L., Holcombe, A. O., & Pashler, H. (2004). Repetition priming in visual search: episodic retrieval, not feature priming. *Memory & Cognition*, 32(1), 12–20.

Hummel, J. E., & Stanikiewicz, B. J. V. C. (1998). Two roles for attention in shape perception: A structural description model of visual scrutiny. *Visual Cognition*, (1–2), 49–79.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506.

Julesz, B. (1981). A theory of preattentive texture discrimination based on first order statistics of textons. *Biological Cybernetics*, *41*, 131–138.

——. (1984). A brief outline of the texton theory of human vision. *Trends in Neuroscience*, *7*(February), 41–45.

Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 179–219.

Klein, R., & Farrell, M. (1989). Search performance without eye movements. *Perception and Psychophysics*, *46*, 476–482.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*(13), 1897–1916.

Kristjansson, A. (2000). In search of rememberance: Evidence for memory in visual search. *Psychological Science*, *11*(4), 328–332.

Lamy, D., & Egeth, H. E. (2003). Attentional capture in singleton-detection and feature-search modes. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(5), 1003–1020.

Li, Z. (2002). A salience map in primary visual cortex. *Trends in Cognitive Sciences*, *6*(1), 9–16.

Luce, R. D. (1986). *Response times*. New York: Oxford University Press.

Luck, S. J., & Vecera, S. P. (2002). Attention. In H. Pashler & S. Yantis (Eds.), *Stevens' handbook of experimental psychology: Vol. 1. Sensation and perception* (3rd ed., pp. 235–286). New York: Wiley.

——, Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *382*, 616–618.

Maioli, C., Benaglio, I., Siri, S., Sosta, K., & Cappa, S. (2001). The integration of parallel and serial processing mechanisms in visual search: evidence from eye movement recording. *European Journal of Neuroscience*, *13*(2), 364–372.

Maljkovic, V., & Nakayama, K. (1994). Priming of popout: I. Role of features. *Memory and Cognition*, 22(6), 657–672.

McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1517–1539.

McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron*, *42*(4), 677–686.

Melcher, D., Papathomas, T. V., & Vidnyánszky, Z. (2005). Implicit attentional selection of bound visual features. *Neuron*, *46*, 723–729.

Mitsudo, H. (2002). Information regarding structure and lightness based on phenomenal transparency influences the efficiency of visual search. *Perception*, *31*(1), 53–66.

Moore, C. M., Egeth, H., Berglan, L. R., & Luck, S. J. (1996). Are attentional dwell times inconsistent with serial visual search? *Psychonomic Bulletin & Review*, 3(3), 360–365.

——, & Wolfe, J. M. (2001). Getting beyond the serial/parallel debate in visual search: A hybrid approach. In K. Shapiro (Ed.), *The limits of attention: Temporal constraints on human information processing* (pp. 178–198). Oxford: Oxford University Press.

Moore, T., Armstrong, K. M., & Fallah, M. (2003). Visuomotor origins of covert spatial attention. *Neuron*, *40*(4), 671–683.

Moraglia, G. (1989). Display organization and the detection of horizontal lines segments. *Perception and Psychophysics*, *45*, 265–272.

Moray, N. (1969). *Attention: Selective processing in vision and hearing*. London: Hutchinson.

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research*, 38(12), 1805–1815.

Murdock, B. B., Jr., Hockley, W. E., & Muter, P. (1977). Two tests of the conveyor-belt model for item recognition. *Canadian Journal of Psychology*, *31*, 71–89.

[AQ3] Najemnik, J., & Geisler, W. S. (in press). Optimal eye movement strategies in visual search. *Nature*.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*, 264–265.

Neisser, U. (1963). Decision time without reaction time: Experiments in visual scanning. *American Journal of Psychology*, *76*, 376–385.

Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research*, *31*(6), 1073–1078.

——. (1992). Feature analysis and the role of similarity in pre-attentive vision. *Perception and Psychophysics*, 52(4), 355–375.

——. (1993). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*, 33(14), 1937–1958.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175.

——, Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). *Top-down control of visual attention in object detection*. Paper presented at the Proceedings of the IEEE International Conference on Image Processing, September 14–17, Barcelona, Spain.

Olzak, L. A., & Thomas, J. P. (1986). Seeing spatial patterns. In K. R. Boff, L. Kaufmann, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Chap. 7). New York: Wiley. [AQ4]

Palmer, J. (1994). Set-size effects in visual search: the effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*(13), 1703–1721.

——. (1995). Attention in visual search: Distinguishing four causes of a set size effect. *Current Directions in Psychological Science*, *4*(4), 118–123.

——. (1998). Attentional effects in visual search: relating accuracy and search time. In R. D. Wright (Ed.), *Visual attention* (pp. 295–306). New York: Oxford University Press.

——, & McLean, J. (1995). *Imperfect, unlimited-capacity, parallel search yields large set-size effects.* Paper presented at the Society for Mathematical Psychology, Irvine, CA.

——, Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*(10–12), 1227–1268.

Pashler, H. E. (1998a). *Attention*. Hove, East Sussex, UK: Psychology Press.

Pashler, H. (1998b). *The psychology of attention*. Cambridge, MA: MIT Press.

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science*, *12*(4), 287–292.

Quinlan, P. T., & Humphreys, G. W. (1987). Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. *Perception and Psychophysics*, *41*, 455–472.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

——, Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159–182.

Rauschenberger, R. (2003). Attentional capture by auto-and allo-cues. *Psychonomic Bulletin & Review*, *10*(4), 814–842.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860.

Rensink, R. A., & Enns, J. T. (1995). Pre-emption effects in visual search: evidence for low-level grouping. *Psychological Review*, *102*(1), 101–130.

——, & Enns, J. T. (1998). Early completion of occluded objects. *Vision Research*, 38, 2489–2505.

——, O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.

Rosenholtz, R. (2001a). Search asymmetries? What search asymmetries? *Perception and Psychophysics*, *63*(3), 476–489.

——. (2001b). Visual search for orientation among heterogeneous distractors: experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance*, 27(4), 985–999.

Sanders, A. F., & Houtmans, M. J. M. (1985). Perceptual modes in the functional visual field. *Acta Psychologica*, *58*, 251–261.

Schall, J., & Thompson, K. (1999). Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience*, *22*, 241–259.

Shapiro, K. L. (1994). The attentional blink: The brain's eyeblink. *Current Directions in Psychological Science*, 3(3), 86–89.

Shen, J., Reingold, E. M., & Pomplun, M. (2003). Guidance of eye movements during conjunctive visual search: the distractor-ratio effect. *Canadian Journal of Experimental Psychology*, 57(2), 76–96.

Shore, D. I., & Klein, R. M. (2000). On the manifestations of memory in visual search. *Spatial Vision*, *14*(1), 59–75.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.

——, & Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends in Cognitive Sciences*, 9(1), 16–20.

Smith, S. L. (1962). Color coding and visual search. *Journal of Experimental Psychology*, *64*, 434–440.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652–654.

Styles, E. A. (1997). *The psychology of attention*. Hove, East Sussex, UK: Psychology Press.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 799–806.

——, Godijn, R., & Pratt, J. (2004). A new estimation of the duration of attentional dwell time. *Psychonomic Bulletin & Review*, *11*(1), 60–64.

Thompson, K. G., & Bichot, N. P. (2004). A visual salience map in the primate frontal eye field. *Progress in Brain Research*, *147*, 249–262.

Thornton, T. (2002). *Attentional limitation and multiple-target visual search*. Unpublished doctoral dissertation, University of Texas at Austin.

Todd, S., & Kramer, A. F. (1994). Attentional misguidance in visual search. *Perception and Psychophysics*, *56*(2), 198–210.

Townsend, J. T. (1971). A note on the identification of parallel and serial processes. *Perception and Psychophysics*, 10, 161–163.

——. (1990). Serial and parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, *1*, 46–54.

Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review*, *11*(3), 391–418.

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*, 156–177.

——. (1986). Features and objects in visual processing. *Scientific American*, *255*(November), 114B–125.

——. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.

——, & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

——, & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15–48.

——, & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 459–478.

——, & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of seperable features. *Journal of Experimental Psychology: General*, *114*, 285–310.

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*, 107–141.

Van Zandt, T. (2002). Analysis of response time distributions. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology*: *Vol. 4. Methodology in experimental psychology* (3rd ed., pp. 461–516). New York: Wiley.

——, & Ratcliff, R. (2005). Statistical mimicking of reaction time data: Single-process models, parameter variability, and mixtures. *Psychonomic Bulletin & Review*, *2*(1), 20–54.

Verghese, P. (2001). Visual search and attention: A signal detection approach. *Neuron*, *31*, 523–535.

von der Malsburg, C. (1981). *The correlation theory of brain function*. Göttingen, Germany: Max-Planck-Institute for Biophysical Chemistry.

Wang, D. L. (1999). Object selection based on oscillatory correlation. *Neural Networks*, *12*(4–5), 579–592.

Ward, R., Duncan, J., & Shapiro, K. (1996). The slow time-course of visual attention. *Cognitive Psychology*, *30*(1), 79–109.

——. (1997). Effects of similarity, difficulty, and nontarget presentation on the time course of visual attention. *Perception and Psychophysics*, *59*(4), 593–600.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

——. (1998). What do 1,000,000 trials tell us about visual search? *Psychological Science*, 9(1), 33–39.

——. (2001). Asymmetries in visual search: An Introduction. *Perception and Psychophysics*, *63*(3), 381–389.

——. (2002). Visual Search: Are some enduring controversies moving toward solution? *Trends in Cognitive Sciences*. Manuscript submitted for publication. **[AQ5]**

——. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences*, *7*(2), 70–76.

Wolfe, J., Alvarez, G., & Horowitz, T. (2000). Attention is fast but volition is slow. *Nature*, *406*, 691.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, 37(1), 25–43.

——, Birnkrant, R. S., Horowitz, T. S., & Kunar, M. A. (2005). Visual search for transparency and opacity:

Attentional guidance by cue combination? *Journal of Vision*, *5*(3), 257–274.

——, Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance*, 29(2), 483–502.

——, Cave, K. R., & Franzel, S. L. (1989). Guided Search: An alternative to the Feature Integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433.

——, & DiMase, J. S. (2003). Do intersections serve as basic features in visual search? *Perception*, 32(6), 645–656.

——, Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 34–49.

——, & Gancarz, G. (1996). Guided Search 3.0: A model of visual search catches up with Jay Enoch 40 years later. In V. Lakshminarayanan (Ed.), *Basic and clinical applications of vision science* (pp. 189–192). Dordrecht, Netherlands: Kluwer Academic.

——, & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501.

——, Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, *435*, 439–440.

[**AQ6**] ——, Michod, K. O., & Horowitz, T. S. (2004). Visual search: Linear reaction time by set size functions are not really linear. *Perception and Psychophysics*. Manuscript submitted for publication.

——, O'Neill, P. E., & Bennett, S. C. (1997). Why are there eccentricity effects in visual search? *Perception and Psychophysics*, in press. [**AQ7**]

——, Palmer, E. M., Horowitz, T. S., & Michod, K. O. (2004). Visual search throws us a curve. *Abstracts of the Psychonomic Society*, *9*. (Paper presented at the meeting of the Psychonomic Society, Minneapolis, MN.)

——, Reinecke, A., & Brawn, P. (2006). Why don't we see changes? The role of attentional bottlenecks and limited visual memory. *Visual Cognition*, in press. [**AQ8**]

——, Yu, K. P., Stewart, M. I., Shorter, A. D., Friedman-Hill, S. R., & Cave, K. R. (1990). Limitations on the parallel guidance of visual search: Color X color and orientation X orientation conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 879–892.

Yantis, S. (1998). Control of visual attention. In H. Pashler (Ed.), *Attention* (pp. 223–256). Hove, East Sussex, UK: Psychology Press.

Zelinsky, G., & Sheinberg, D. (1995). Why some search tasks take longer than others: Using eye movements to redefine reaction times. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications: Vol. 6. Studies in visual information processing* (pp. 325–336). Amsterdam, Netherlands: Elsevier.

Zelinsky, G. J., & Sheinberg, D. L. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research*, 36(14), 2177–2187.

——. (1997). Eye movements during parallel/serial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1), 244–262.

Zohary, E., Hochstein, S., & Hillman, P. (1988 ). Parallel and serial processing in detecting conjunctions. *Perception*, *17*, 416.

**Author Query Sheet**

Manuscript No.: Chapter-08

Dear Author,

The following queries have arisen during the setting of your manuscript. Please answer the queries using the form below.

| Query no. | Query | Reply |
|---|---|---|
| AQ1 | Should the asterisk be changed to a multiplication sign in “2*deg”? | |
| AQ2 | Should y-axis be the RT(reaction time)? If so, add label “RT” (or Reaction Time). | |
| AQ3 | Update publication information for Najemnik & Geisler (in press). | |
| AQ4 | Please provide page numbers of chapter. | |
| AQ5 | Update publication information for Wolfe (2002). | |
| AQ6 | Update publication information for Wolfe, Michod, & Horowitz (2004, submitted). | |
| AQ7 | Update publication information for Wolfe, O’Neill, & Bennett (1997, in press). | |
| AQ8 | Update publication information for Wolfe, Reinecke, & Brawn (2006, in press). | |